# Lecture 8: Generalized Least Squares

Prepared for ECON 5033

Eric S. Lin

Department of Economics, National Tsing Hua University

November 19, 2024

# Outline

- General non-spherical errors – relaxing homoskedasticity and non-serial correlation
- Generalized least squares approach
- Heteroskedasticity

# Basic Assumptions

- Let's reconsider the classical assumptions:
  - Zero covariance
  - Constant variance across observations
- Linear regression model becomes "generalized" linear regression (GLR) model.
- Recall the assumptions we have before:
  - The $n \times k$ data matrix $X$ is of full column rank, i.e., $\text{rank}(X) = k$ and $n > k$.
  - Conditional on $X$, the mean of $Y$ is a linear combination of the columns of $X$: $\mathrm{E}[Y|X] = X\beta$.
  - Conditional on $X$, the variance covariance matrix of $Y$ is $\text{var}[Y|X] = \sigma^2 I_n$.

# Basic Assumptions

- In GLR model, we say that the conditional variance is no longer a sphere.
- Conditional on $X$, the variance covariance matrix of $Y$ is $\text{var}[Y|X] = \sigma^2\Omega \neq \sigma^2 I$, where $\Omega$ is symmetric and positive definite.
- The non-spherical disturbances include heteroskedasticity and autocorrelation.
- Heteroskedasticity:
  - For the $i$th error term: $\text{E}[\varepsilon_i^2] = \sigma_i^2 = \sigma^2(X_i) \neq \sigma^2$
  - Variance-covariance matrix of $\varepsilon$
- Autocorrelation
  - For the $i$ and $j$th error terms: $\text{cov}[\varepsilon_i, \varepsilon_j] \neq 0$
  - Variance-covariance matrix of $\varepsilon$

# Autocorrelation

## Example

Consider the stationary autoregressive of order one, $AR(1)$, process:

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t, \ \ |\rho| < 1, \ v_t \overset{i.i.d.}{\sim} (0, \sigma_v^2).$$

Using the fact that $\varepsilon_t = \sum_{j=0}^{\infty} \rho^j v_{t-j}$ leads to the following:

$$\text{var}[\varepsilon_t] = \sum_{j=0}^{\infty} \rho^{2j} \text{var}[v_{t-j}] = \frac{\sigma_v^2}{1-\rho^2}$$

$$\text{cov}[\varepsilon_t, \varepsilon_{t-s}] = \text{cov}[\sum_{j=0}^{\infty} \rho^j v_{t-j}, \sum_{j=0}^{\infty} \rho^j v_{t-s-j}] = \frac{\sigma_v^2 \rho^{|s|}}{1-\rho^2}.$$

# Autocorrelation

## Example

Consider the following static one-factor random component model (RE Panel model):

$$y_{it} = x_{it}'\beta + \varepsilon_{it}, \;\; i = 1, ..., N; \;\; t = 1, ..., T$$
$$\varepsilon_{it} = \alpha_i + u_{it},$$

where $\alpha_i$ are i.i.d. $(0, \sigma_\alpha^2)$ and independent of $u_{it}$ and $x_{it}$ $\forall i, t$. $u_{it}$ are i.i.d. $(0, \sigma_u^2)$. To get a full understanding of the variance covariance structure, we first consider the following:

$$E[\varepsilon_i \varepsilon_i'] = E[(\iota_T \alpha_i + u_i)(\iota_T \alpha_i + u_i)'] = \sigma_\alpha^2 \iota_T \iota_T' + \sigma_u^2 I_T$$
$$E[\varepsilon_i \varepsilon_j'] = E[(\iota_T \alpha_i + u_i)(\iota_T \alpha_j + u_j)'] = 0$$

# Autocorrelation

## Example

The variance covariance structure of $\varepsilon$ is

$$E[\varepsilon\varepsilon'] = \sigma_u^2 I_{NT} + \sigma_\alpha^2 (I_N \otimes \iota_T \iota_T').$$

One can see that $\varepsilon_{it}$ is homoskedastic but serial correlated.

1. Off-diagonal blocks ($i \neq j$)
2. Diagonal blocks ($i = j$): diagonal and off-diagonal elements

# OLS under Nonspherical Disturbance

- The regression model is $Y = X\beta + \varepsilon$ with $\text{var}[\varepsilon] = \sigma^2\Omega$. The OLS estimator is,

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y.$$

- Also, one can show,

$$\begin{aligned} \mathsf{E}[\hat{\beta}] &= \beta \\ \text{var}[\hat{\beta}] &= \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1} \neq \sigma^2(X'X)^{-1}. \end{aligned}$$

  - Problem using incorrect standard errors?

- Under the Gaussian errors, we have,

$$\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}).$$

  - Using the new (correct) standard errors to do inference.

# OLS under Nonspherical Disturbance

- If the errors have enough moments and there is not too much serial correlation, one would be able to establish the consistency and asymptotic normality results for OLS estimator.

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2 Q^{-1} \text{plim}(\frac{X'\Omega X}{n}) Q^{-1}),$$

where $Q = \text{plim}(X'X/n)$ .

- OLS estimator, however, is not BLUE in non-spherical disturbances.
- Better estimator than OLS?
    - GLS turns out to be the better estimator.

# Too Much Serial Correlation

## Example

Here is an example of too much serial correlation.

$$\sigma^2\Omega = \sigma^2 \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \dots & \dots & \dots & \dots \\ \rho & \rho & \dots & 1 \end{bmatrix}.$$

# GLS Transformation

- Before introducing GLS transformation, we provide a useful lemma.

## Lemma

*Given $\Omega$ a positive definite matrix, there exists a square, nonsingular matrix $\omega$ which satisfies $\omega'\omega = \Omega^{-1}$.*

## Proof.

If $\Omega$ is a positive definite matrix, there exist $C$ and $\Lambda$ such that

$$\Omega = C\Lambda C' \quad \text{with } C'C = I.$$

$$\Omega = C\Lambda^{1/2}\Lambda^{1/2}C'$$

$$\Omega^{-1} = (C')^{-1}\Lambda^{-1/2}\Lambda^{-1/2}C^{-1}$$

$$\Omega^{-1} = C\Lambda^{-1/2}\Lambda^{-1/2}C' = \omega'\omega.$$

$\square$

# GLS Transformation

- Let's pre-multiply the regression model by $\omega$ defined above,

$$\omega Y = \omega X\beta + \omega\varepsilon$$
$$Y^* = X^*\beta + \varepsilon^*.$$

- The variance of transformed error term is,

$$\text{var}[\varepsilon^*|X^*] = \text{E}[\varepsilon^*\varepsilon^{*\prime}|X^*] = \sigma^2 I$$

- The transformed model satisfies the assumption of spherical errors.
- Now we could apply the Gauss-Markov Theorem that OLS on the transformed model is BLUE.

# GLS Estimator

- The GLS estimator (or Aitken estimator) is,

$$\hat{\beta}_{GLS} = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^* = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y.$$

- One can show the following properties of GLS estimator.

$$\begin{aligned} \mathrm{E}[\hat{\beta}_{GLS}] &= \beta + \mathrm{E}[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon] = \beta \\ \mathrm{var}[\hat{\beta}_{GLS}] &= \sigma^2(X'\Omega^{-1}X)^{-1} \end{aligned}$$

- With normality assumption, $\varepsilon \sim \mathcal{N}(0, \sigma^2\Omega)$, it is easy to show,

$$\hat{\beta}_{GLS} \sim \mathcal{N}(\beta, \sigma^2(X'\Omega^{-1}X)^{-1}).$$

- Once again, the Gauss-Markov theorem could be applied to get efficiency of GLS estimator.
- Can we show it directly?

# Efficient GLS Estimator

**Theorem**

*GLS estimator is more efficient than OLS estimator.*

**Proof.**

One could show,

$$
\begin{aligned}
& X'\Omega^{-1}X - X'X\left(X'\Omega X\right)^{-1}X'X \\
& = X'\Omega^{-1/2}(I - \Omega^{1/2}X(X'\Omega X)^{-1}X'\Omega^{1/2})\Omega^{-1/2}X,
\end{aligned}
$$

which is p.s.d. $\square$

# Large Sample Properties of GLS Estimator

- Without imposing normality assumption, we can still yield some nice large sample properties of GLS estimators.
- Provided that the transformed $X^*$ have sufficient moments then $\text{plim}\hat{\beta}_{GLS} = \beta$ or $\hat{\beta}_{GLS} \xrightarrow{p} \beta$.
- We can also have that:

$$\sqrt{n}(\hat{\beta}_{GLS} - \beta) \xrightarrow{d} \mathcal{N}(0, \sigma^2 \text{plim}(\frac{X'\Omega X}{n})^{-1}).$$

# GLS vs. MLE

- All tests go through as usual in terms of the transformed model and GLS is the MLE under the assumption that $\varepsilon \sim \mathcal{N}(0, \sigma^2\Omega)$.
- The log-likelihood function has the form,

$$\ln \mathcal{L} = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(|\sigma^2\Omega|) - \frac{1}{2\sigma^2}(Y - X\beta)'\Omega^{-1}(Y - X\beta).$$

- With $\Omega$ a matrix of known constants, the MLE of $\beta$ is the vector that minimizes the generalized residual sum of squares.

$$GRSS(\beta) = (Y - X\beta)'\Omega^{-1}(Y - X\beta)$$

- One will yield,

$$\hat{\beta}_{MLE} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y = \hat{\beta}_{GLS}$$
$$\hat{\sigma}^2_{MLE} = \frac{1}{n}(Y - X\hat{\beta}_{MLE})'\Omega^{-1}(Y - X\hat{\beta}_{MLE}).$$

# Some Pitfalls

- For OLS:
  - You may use wrong standard errors, i.e., $\hat{\sigma}^2(X'X)^{-1}$, even $\hat{\beta}$ is still consistent
    - Inconsistent standard errors in plim sense.
    - Inconsistent estimator of error variance (non-constant now!)
    - Misleading inference, i.e., all tests are invalid.
  - Even if you happen to get the right standard error, it's not good enough.
    - $\text{var}[\hat{\beta}] = \sigma^2(X'X)^{-1}X'\Omega X(X'X)^{-1}$
    - Efficiency issue!
- For GLS:
  - Typically $\Omega$ is unknown.
    - It's known as infeasible GLS.
  - Will need to do feasible GLS or FGLS later on.

# Why Is Infeasible GLS?

- What is the dimension of $\Omega$?
- The number of parameters to estimate is larger than $n$ and increases faster than $n$.
- There is no way to devise an estimator that satisfies a LLN without adding restrictions.
- The feasible GLS estimator is based on making sufficient assumptions regarding the form of $\Omega$ so that a consistent estimator can be devised.

# How to Do Feasible GLS?

- Suppose that we parameterize $\Omega$ as a function of $X$ and $\theta$, where $\theta$ may include $\beta$ as well as other parameters, so that:

$$\Omega = \Omega(X, \theta),$$

where $\theta$ is of fixed dimension.

- If we can consistently estimate $\theta$, we can consistently estimate $\Omega$, as long as $\Omega(X, \theta)$ is a continuous function of $\theta$ (by Slutsky theorem).
- In this case,

$$\hat{\Omega} = \Omega(X, \hat{\theta}) \xrightarrow{p} \Omega(X, \theta)$$

- If we replace $\Omega$ in the formula for the GLS estimator with $\hat{\Omega}$, we obtain the FGLS estimator.

# How to Do Feasible GLS?

- The FGLS estimator can be obtained as follows:

$$\hat{\omega} Y = \hat{\omega} X \beta + \hat{\omega} \varepsilon,$$

  where $\hat{\omega} = \texttt{Chol}(\hat{\Omega}^{-1})$ – Cholesky decomposition

- The FGLS estimator shares the same asymptotic properties as GLS:
  1. Consistency
  2. Asymptotic normality
  3. Asymptotic efficiency if the errors are normally distributed (CRLB)
  4. Test procedures are asymptotically valid

# Midterm #2

- Midterm #2 will be held on November 26 from 9 to noon
- Coverage includes:
  1. multiple regression model
     FWL, testing, RLS
  2. asmptotic theory
     consistency, LLN, CLT
  3. trinity of tests
     W, LR, LM
  4. Chow test, dummies, collinearity
  5. GLS basics
- Two A4 formula sheets are allowed